

DOCUMENT RESUME

ED 063 334

TM 001 352

AUTHOR Dziuban, Charles D.; Harris, Chester W.
TITLE On the Extraction of Components and the Applicability
of the Factor Model.
PUB DATE Apr 72
NOTE 12p.; Paper presented at the annual meeting of the
American Educational Research Association (Chicago,
Illinois, April 1972)
EDRS PRICE MF-\$0.65 HC-\$3.29
DESCRIPTORS *Correlation; *Factor Analysis; *Mathematical Models;
Predictor Variables; *Psychometrics; *Statistical
Analysis
IDENTIFIERS Project TALENT; Shaycroft (M F)

ABSTRACT

A reanalysis of Shaycroft's matrix of intercorrelations of 10 test variables plus 4 random variables is discussed. Three different procedures were used in the reanalysis: (1) Image Component Analysis, (2) Uniqueness Rescaling Factor Analysis, and (3) Alpha Factor Analysis. The results of these analyses are presented in tables. It is concluded from the results that factor analysis or image component analysis offers protection from interpreting random relationships as being meaningful.
(Author/DB)

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRO-
DUCED EXACTLY AS RECEIVED FROM
THE PERSON OR ORGANIZATION ORIG-
INATING IT. POINTS OF VIEW OR OPIN-
IONS STATED DO NOT NECESSARILY
REPRESENT OFFICIAL OFFICE OF EDU-
CATION POSITION OR POLICY.

ON THE EXTRACTION OF COMPONENTS AND THE APPLICABILITY
OF THE FACTOR MODEL

by

Charles D. Dziuban
Florida Technological University

and

Chester W. Harris
University of California, Santa Barbara

Presented at the annual meeting of
The American Educational Research Association

Chicago
April 3-7, 1972

ON THE EXTRACTION OF COMPONENTS
AND THE APPLICABILITY OF THE FACTOR MODEL

It is well known that the routine use of principal component analysis with correlation matrices may cause problems in the interpretation of results. Armstrong and Soelberg (1968) illustrated how one might be led to interpret components of intercorrelations among random normal deviates, the expected values of which were zero. As suggested by Kaiser (1960), they rotated only those components with corresponding eigenvalues greater than one and obtained an ostensibly good structure. The authors pointed out, however, that any interpretation of "meaningfulness" of these components would be erroneous and recommended that measures of reliability accompany principal component procedures.

Tobias and Carlson (1969) demonstrated that such inappropriate analyses can be guarded against by prior application of Bartlett's Test of Sphericity. The test statistic is computed by the formula:
$$- ((N-1) - 1/6 (2P + 5)) \log e |R|$$
where N is the sample size, P is the number of variables and $|R|$ is the determinant of the correlation matrix. For large N the statistic is approximately distributed as chi square with $1/2 (p - 1)$ degrees of freedom and has the associated hypothesis that the sample correlation matrix came from a multivariate normal population in which the variables of interest are independent. Tobias and Carlson argued that failure to reject should preclude analysis via principal components and illustrated that the probability was .55 that the matrix analyzed by Armstrong and Soelberg was a sample from an uncorrelated population. They

recommended that correlation matrices not be submitted to component analysis when the matrix yields nonsignificant results on Bartlett's test.

Recently, Shaycroft (1970) presented a principal component analysis of a matrix of correlations among ten tests of interest plus four random deviates. Once again, the criterion for component retention was the number of eigenvalues of R greater than one, a procedure which yielded three components. These are reprinted in Table I. The first component is correlated substantially with ten tests of interest, the second and third components are correlated primarily with two of the random deviates. This result is understandable, of course, since the first component is located near the centroid of the ten tests and the subsequent components are orthogonal to this first component and thus in a position near the random variables that are essentially uncorrelated with the ten tests.

We note that the application of Bartlett's test to Shaycroft's matrix would not have guarded against the component analysis since the intercorrelations of the ten variables of interest were not randomly different from zero. (The determinant of the complete correlation matrix was quite small.) These variables were in fact tests used in connection with project TALENT. Shaycroft's results, taken at face value, strongly suggest that a random variable be interpreted as the core of a "meaningful" component. The question of how to guard against a result like Shaycroft's therefore arises.

The answer to this question is relatively simple. Whenever one has reason to suspect that one or more of the variables being analyzed are

operating like random variables, essentially uncorrelated with the remaining variables, one should abandon Little Jiffy and turn to another model. This advice is not inconsistent with a position taken by Kaiser (1970) recently. To illustrate, we reanalyzed Shaycroft's matrix of intercorrelations of ten test variables plus four random variables using three different procedures:

1) Image Component Analysis (Guttman, 1953) which is based on the image covariance matrix $R + S^2 R^{-1} S^2 - 2S^2$. Components were retained according to the number of eigenvalues of $S^{-1}RS^{-1}$ greater than one -- the strong lower bound.

2) Uniqueness Rescaling Factor Analysis (Harris, 1962) which is based on the correlation matrix after it has been rescaled with uniqueness estimates to give $U^{-1}RU^{-1}$. Squared multiple correlations subtracted from one were used initially to estimate uniquenesses. Factors were retained according to the number of eigenvalues of the matrix greater than one -- also the strong lower bound.

3) Alpha Factor Analysis (Kaiser and Caffrey, 1965) which is based on the correlation matrix reduced with uniqueness and rescaled with communality estimates $H^{-1}(R-U^2)H^{-1}$. The method is iterative and produces factors with maximum generalizability in the sense of Cronbach's alpha. Factors were retained according to the eigenvalues of the matrix greater than one -- the weak lower bound.

All raw pattern coefficients were orthogonally rotated according to the normal varimax criterion. The results of these analyses are presented

in Table II (A, B, C).

The image procedure extracted seven components while uniqueness rescaling and alpha retained seven and four factors respectively. It can be observed that the correlations of the random variables with the image components or with the two sets of factors did not warrant interpretation. The highest correlations were $-.038$ and $+.038$ for X_1 and X_4 on components three and five for the image solution. The highest correlation for the uniqueness rescaling procedure was $+.141$ for X_4 on the fifth factor. The highest coefficient for the alpha procedure was $+.171$ for X_4 on the second factor. These results seem reasonable when the uniqueness estimates of the variables presented in Table III are examined. The estimates for each random deviate approached unity which was to be expected since any exhibited correlation differed only randomly from zero.

Some time ago Harris (1964) pointed out again that principal component analysis is a descriptive procedure yielding uncorrelated derived variables within the variable space. Extracting components of R corresponding to eigenvalues greater than unity is not a factor analysis in the Spearman/Thurstone sense. Armstrong and Soelberg, and Shaycroft have given examples of difficulties that can arise from using Little Jiffy. We have shown here that not all contingencies can be guarded against by routine use of the Bartlett test. Our point is that it is factor analysis or image component analysis that offers protection from interpreting random relationships as being meaningful.

REFERENCES

- Armstrong, J. S. and P. Soelberg, "On the Interpretation of Factor Analysis," Psychological Bulletin, 1968, 70, 361-364.
- Guttman, L., "Image Theory for the Structure of Quantitative Variates," Psychometrika, 1953, 18, 277-296.
- Harris, C. W., "Some Rao Guttman Relationships," Psychometrika, 1962, 27, 247-263.
- Harris, C. W., "Some Recent Developments in Factor Analysis," Educational and Psychological Measurement, 1964, 24, 193-205.
- Kaiser, H. F., "The Application of Electronic Computers to Factor Analysis," Educational and Psychological Measurement, 1960, 20, 141-151.
- Kaiser, H. F., "A Second-Generation Little Jiffy," Psychometrika, 1970, 35, 401-416.
- Kaiser, H. F. and John Caffrey, "Alpha Factor Analysis," Psychometrika, 1965, 30, 1-14.
- Shaycroft, M. F., "The Eigenvalue Myth and the Dimension-Reduction Fallacy," (Paper Presented at the American Educational Research Association, Minneapolis, Minnesota, March 6, 1970).
- Tobias, Sigmund and J. E. Carlson, "Brief Report: Bartlett's Test of Sphericity and Chance Findings in Factor Analysis," Multivariate Behavioral Research, 1969, 4, 375-377.

TABLE I
DERIVED PRINCIPAL COMPONENTS (NORMAL VARIMAX)
(Reprinted from Shaycroft)

		VARIMAX LOADINGS			
		V1	V2	V3	$\sum b^2$ *
1	R-102	.859	-.015	-.039	.740
2	R-103	.794	.020	-.065	.635
3	R-106	.865	.006	-.004	.748
4	R-107	.854	-.028	-.016	.730
5	R-112	.595	-.041	.019	.356
6	R-230	.766	.034	-.050	.590
7	R-250	.839	.037	-.041	.707
8	R-282	.605	.015	.095	.375
9	R-290	.711	.034	.077	.513
10	R-312	.830	.009	.015	.689
11	X-1	-.025	-.681	-.060	.468
12	X-2	.029	-.180	.721	.553
13	X-3	-.023	.190	.686	.507
14	X-4	-.004	.704	-.051	.498
$\sum b^2$ **		6.053	1.035	1.022	

* Sums of the Squared Row Elements

** Sums of the Squared Column Elements

TABLE II A

DERIVED IMAGE COMPONENTS (NORMAL VARIMAX)

	I	II	III	IV	V	VI	VII
R-102 Vocab.	.639	.391	.145	.170	-.104	.051	.033
R-103 Lit. Info.	.591	.426	.158	.149	.147	.084	.031
R-106 Math Info.	.559	.400	.360	.391	.001	.033	.038
R-107 Phys. Sci. Info.	.636	.379	.204	.296	.028	.019	.043
R-112 Mech. Info.	.557	.156	.004	.095	.007	-.029	.023
R-230 English	.516	.332	.332	.148	.111	.089	-.005
R-250 Rdg. Comp.	.652	.341	.257	.108	.142	.077	-.007
R-282 Vis. 3 Dimen.	.514	.012	.266	.140	-.045	-.048	-.044
R-290 Abst. Reas.	.553	.110	.254	.152	.013	-.000	-.049
R-312 Intro. H.S. Math	.502	.366	.426	.379	-.002	.043	.023
X1 Random	-.015	.004	-.038	-.006	-.023	-.000	-.001
X2 Random	.011	-.033	.005	.037	-.009	.000	-.000
X3 Random	-.013	-.033	.002	.004	.001	.001	.000
X4 Random	.002	.001	.006	-.008	.038	.000	-.000
Component Variance	3.372	1.037	.772	.522	.070	.030	.011
% Component Variance	58.000	17.800	13.300	9.000	1.200	.500	.200
% Total Variance	24.100	7.400	5.500	3.700	.500	.200	.100
Eigenvalues	18.721	2.199	1.746	1.500	1.042	1.017	1.016

TABLE II B

DERIVED UNIQUENESS RESCALING FACTORS (NORMAL VARIMAX)

	I	II	III	IV	V	VI	VII
R-102 Vocab.	.800	.168	.308	-.050	.011	-.022	.025
R-103 Lit. Info.	.795	.055	.174	-.090	.067	-.040	.053
R-106 Math Info.	.818	.247	.010	.294	.005	.085	.059
R-107 Phys. Sci. Info.	.777	.2071	.242	.159	-.021	.056	.077
R-112 Mech. Info.	.431	.244	.452	-.012	-.061	.016	-.001
R-230 English	.724	.235	-.063	-.058	.077	-.059	-.068
R-250 Rdg. Comp.	.772	.265	.122	-.159	.095	-.058	-.011
R-282 Vis. 3 Dimen.	.354	.579	.153	.043	.021	.065	-.006
R-290 Abst. Reas.	.511	.551	.032	-.010	.073	.039	-.012
R-312 Intro. H.S. Math	.780	.289	-.113	.289	.020	.075	.014
X1 Random	-.018	-.024	.010	-.014	-.123	-.008	.030
X2 Random	.016	.006	.001	.007	-.008	.115	.019
X3 Random	-.029	.006	.001	-.008	.043	.097	-.007
X4 Random	.004	-.014	.005	-.014	.141	.007	.023
Factor Variance	4.846	1.042	.445	.237	.067	.054	.022
% Factor Variance	72.200	15.500	6.600	3.500	1.000	.800	.300
% Total Variance	34.600	7.400	3.200	1.700	.500	.400	.200
Eigenvalues	18.721	2.199	1.746	1.500	1.042	1.017	1.016

TABLE II C
DERIVED ALPHA FACTORS (NORMAL VARIMAX)

	I	II	III
R-102 Vocab.	.852	-.020	-.046
R-103 Lit. Info.	.747	.068	-.070
R-106 Math Info.	.845	.023	.027
R-107 Phys. Sci. Info.	.845	-.061	.001
R-112 Mech. Info.	.548	-.025	.027
R-230 English	.721	.079	-.094
R-250 Rdg. Comp/	.821	.109	-.083
R-282 Vis. 3 Dimen.	.573	.030	.064
R-290 Abst. Reas.	.685	.059	.100
R-321 Intro. H.S. Math	.799	.017	.073
X1 Random	-.023	-.157	-.015
X2 Random	.019	-.023	.132
X3 Random	-.022	.050	.142
X4 Random	-.002	.171	-.001
Factor Variance	5.644	.090	.081
% Factor Variance	97.100	1.500	1.400
% Total Variance	40.300	.600	.600
Eigenvalues	9.944	2.402	1.897

TABLE III
 s_j^2 UNIQUENESS ESTIMATES FOR THE VARIABLES

<u>Variable</u>	<u>s_j^2</u>
R-102 Vocab.	.282
R-103 Lit. Info.	.371
R-106 Math Info.	.214
R-107 Phys. Sci. Info.	.305
R-112 Mech. Info.	.619
R-230 English	.448
R-250 Rdg. Comp.	.334
R-282 Vis. 3 Dimen.	.600
R-290 Abst. Reas.	.499
R-312 Intro. H.S. Math	.253
X1 Random	.997
X2 Random	.998
X3 Random	.998
X4 Random	.997

TABLE IV

CORRELATION MATRIX
(Reprinted From Shaycroft)

First 10 variables: TALENT Scores
Last 4 variables: Random Numbers

Cases: A representative 10% sample of those 12th-grade boys in
Project TALENT having scores on all 10 of the test variables.

No. of cases: 3689

C O R R E L A T I O N C O E F F I C I E N T S

TALENT test	1	2	3	4	5	6	7	8	9	10	11	12	13	14	$\bar{X}$	s
1 R-102 Vocabulary		.726	.690	.733	.568	.614	.723	.417	.506	.615	-.014	.012	-.022	-.003	14.08	3.80
2 R-103 Literature Info.			.648	.672	.398	.582	.700	.321	.442	.575	-.020	.012	-.024	.020	14.12	4.74
3 R-106 Math Information				.764	.405	.608	.639	.456	.554	.840	-.026	.024	-.021	.004	11.42	6.18
4 R-107 Physical Science Info.					.515	.558	.650	.442	.521	.689	-.008	.022	-.026	-.005	10.44	4.37
5 R-112 Mechanical Information						.345	.441	.400	.341	.349	-.006	.010	-.004	-.003	13.37	3.36
6 R-230 English							.659	.366	.501	.653	-.029	-.003	-.021	.005	82.77	13.34
7 R-250 Reading Comprehension								.440	.577	.616	-.029	.005	-.023	.020	32.99	10.65
8 R-282 Visualization in Three Dimensions									.579	.442	-.026	.015	-.009	-.001	9.75	3.45
9 R-290 Abstract Reasoning										.558	-.027	.018	.004	.002	9.61	3.00
10 R-312 Math 11. Introd. H.S. Math										-.029		.029	-.011	-.006	12.41	5.61
11 X-1 Random numbers*												.001	-.009	-.028	.4922	.2885
12 X-2 Random numbers*													.019	-.005	.4997	.2891
13 X-3 Random numbers*														.010	.5013	.2886
14 X-4 Random numbers*															.5000	.2880

* Four-digit random numbers between .0000 and .9999, with an approximately rectangular distribution.